

Normalizing the causality between time series

X. San Liang*

Nanjing University of Information Science and Technology

(Nanjing Institute of Meteorology), Nanjing 210044, and

China Institute for Advanced Study,

Central University of Finance and Economics, Beijing 100081, China

(Dated: January 23, 2015)

Abstract

Recently, a rigorous yet concise formula has been derived to evaluate the information flow, and hence the causality in a quantitative sense, between time series. To assess the importance of a resulting causality, it needs to be normalized. The normalization is achieved through distinguishing three types of fundamental mechanisms that govern the marginal entropy change of the flow recipient. A normalized or relative flow measures its importance relative to other mechanisms. In analyzing realistic series, both absolute and relative information flows need to be taken into account, since the normalizers for a pair of reverse flows belong to two different entropy balances; it is quite normal that two identical flows may differ a lot in relative importance in their respective balances. We have reproduced these results with several autoregressive models. We have also shown applications to a climate change problem and a financial analysis problem. For the former, reconfirmed is the role of the Indian Ocean Dipole as an uncertainty source to the El Niño prediction. This might partly account for the unpredictability of certain aspects of El Niño that has led to the recent portentous but spurious forecasts of the 2014 “Monster El Niño”. For the latter, an unusually strong one-way causality has been identified from IBM (International Business Machines Corporation) to GE (General Electric Company) in their early era, revealing to us an old story, which has almost gone to oblivion, about “Seven Dwarfs” competing a giant for the mainframe computer market.

PACS numbers: 05.45.-a, 89.70.+c, 89.75.-k, 02.50.-r

Keywords: Causality; Time series; Information flow; Uncertainty propagation; Climate change; Financial economics

*Electronic address: sanliang@courant.nyu.edu

I. INTRODUCTION

Information flow, or information transfer as it may be referred to in the literature, has long been recognized as the appropriate measure of causality between dynamical events[1]. It possesses the needed asymmetry or directionalism for a cause-effect relation, and, moreover, provides a quantitative characterization of the otherwise statistical test, e.g., the Granger causality test[2]. For this reason, the past decades have seen a surge of interest in this arena of research. Measures of information flow proposed thus far include, for example, the time-delayed mutual information[3], transfer entropy[4], momentary information transfer[6], causation entropy[7], etc., among which transfer entropy has been proved to be equivalent to Granger causality up to a factor 2 for linear systems[5].

Recently, Liang and Kleeman find that the notion of information flow actually can be put on a rigorous footing within a given deterministic system[8]. The basic idea can be best illustrated with a system of two components, say, X_1 and X_2 . The problem here essentially deals with how the marginal entropies of X_1 and X_2 , written respectively as H_1 and H_2 , evolve. Take H_1 for an example. Its evolution could be due to X_1 its own and/or caused by X_2 . That is to say, dH_1/dt can be split exclusively into two parts:

$$\frac{dH_1}{dt} = \frac{dH_1^*}{dt} + T_{2 \rightarrow 1},$$

if we write the contribution from the former mechanism as dH_1^*/dt and that from the latter as $T_{2 \rightarrow 1}$. This $T_{2 \rightarrow 1}$ is the very time rate of information flowing from X_2 to X_1 .

To find the information flow $T_{2 \rightarrow 1}$, it suffices to find dH_1^*/dt , since, for each deterministic system, there is a Liouville equation for the density of the state and, accordingly, dH_1/dt can be obtained. In Ref. [8], dH_1^*/dt is acquired through an intuitive argument based on an entropy evolutionary law established therein. The same result is later on rigorously proved; see [9] for a review. For stochastic systems which we will be considering in this study, the trick ceases to work, but in [10] Liang manages to circumvent the difficulty and find the result, which we will be briefly reviewing in the following.

Consider a two-dimensional (2D) stochastic system

$$d\mathbf{X} = \mathbf{F}(\mathbf{X}, t)dt + \mathbf{B}(\mathbf{X}, t)d\mathbf{W}, \quad (1)$$

where $\mathbf{F} = (F_1, F_2)^T$ is the vector of drift coefficients (differentiable vector field), $\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$ the matrix of stochastic perturbation coefficients, and \mathbf{W} a 2D standard Wiener process. Let $g_{ij} = \sum_k b_{ik}b_{jk}$, and ρ_i be the marginal probability density function of x_i . Liang (2008)[10] proves that the time rate of information flowing from X_2 to X_1 is

$$T_{2 \rightarrow 1} = -E \left(\frac{1}{\rho_1} \frac{\partial F_1 \rho_1}{\partial x_1} \right) + \frac{1}{2} E \left(\frac{1}{\rho_1} \frac{\partial^2 g_{11} \rho_1}{\partial x_1^2} \right), \quad (2)$$

where E signifies the operator mathematical expectation. This measure of information flow is asymmetric between the two parties and, particularly, if the process underlying X_1 does not depend on X_2 , then the resulting causality from X_2 to X_1 vanishes. This is the so-called *property of causality*, which asserts, in the above language, that $T_{2 \rightarrow 1}$ vanishes if F_1 and g_{11} are independent of x_2 . When $T_{2 \rightarrow 1}$ is nonzero, it may take positive or negative values. A positive $T_{2 \rightarrow 1}$ means X_2 causes X_1 to be more uncertain, while a negative $T_{2 \rightarrow 1}$ reduces

the entropy of X_1 , and hence functions to stabilize the latter. For more details, referred to Ref. [9].

The above theorem is recently applied to time series analysis. Under the assumption of a linear model with additive noise, Liang[11] shows that the maximal likelihood estimate of the information flow in (2) turns out to be very tight in form, involving only the common statistics namely sample covariances. Take two series X_1 and X_2 for example. The rate of information flowing (units: nats per unit time) from X_2 to X_1 is shown to be

$$T_{2 \rightarrow 1} = \frac{C_{11}C_{12}C_{2,d1} - C_{12}^2C_{1,d1}}{C_{11}^2C_{22} - C_{11}C_{12}^2}, \quad (3)$$

where C_{ij} is the sample covariance between X_i and X_j , and $C_{i,dj}$ is that between X_i and \dot{X}_j , \dot{X}_j being the difference approximation of dX_j/dt using the Euler forward scheme:

$$\dot{X}_j(n) = \frac{X_j(n+k) - X_j(n)}{k\Delta t}. \quad (4)$$

Here k is usually 1, but for highly chaotic and densely sampled series, $k = 2$ should be chosen to avoid getting spuriously large \dot{X}_j due to possible shock structures that make the differencing highly sensitive to the error in X_j . This formula involves only sample covariances, and is hence very convenient to evaluate. In addition, it is easy to see that if $C_{12} = 0$ then $T_{2 \rightarrow 1} = 0$, but when $T_{2 \rightarrow 1} = 0$ the correlation C_{12} does not need to vanish. That is to say, contrapositively, *causation implies correlation, but correlation does not imply causation*. In an explicitly quantitative way, this corollary resolves the long-standing debate over causation versus correlation.

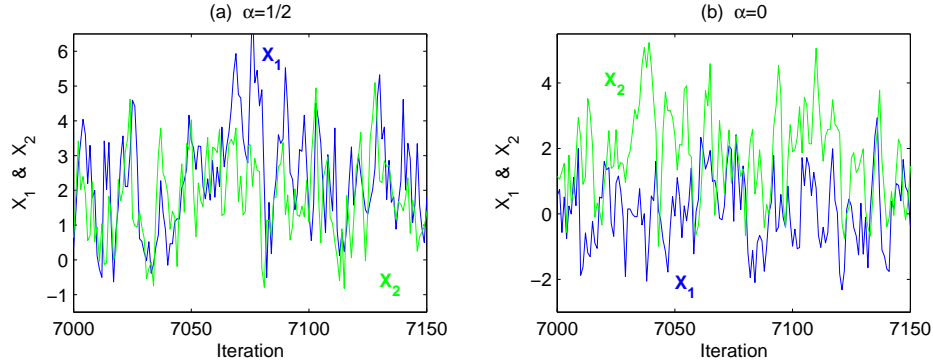


FIG. 1: (Color online) A sample path (X_1 and X_2) of the autoregressive process (5) for (a) $\alpha = 0.5$, $\beta = 0$, and (b) $\alpha = 0$, $\beta = 0$.

In general, the rates of information flow differ from case to case. One would like to normalize them in real applications, just as that with correlation coefficients, in order to assess the importance of the flow identified, if any. For example, consider two series generated from two autoregressive processes:

$$X_1(n+1) = 0.1 + 0.5X_1(n) + \alpha X_2(n) + e_1(n), \quad (5a)$$

$$X_2(n+1) = 0.7 + \beta X_1(n) + 0.6X_2(n) + e_2(n), \quad (5b)$$

where the errors $e_1 \sim N(0, 1)$ and $e_2 \sim N(0, 1)$ are independent. Generate a pair of series with 80000 steps on MATLAB, and perform the causality analysis (with $k = 1$), one obtains (units are in nats per iteration; same below in this section)

(1) for $(\alpha, \beta) = (0.5, 0)$, $T_{2 \rightarrow 1} = 0.1481$, $T_{1 \rightarrow 2} = -0.0002$;

(2) for $(\alpha, \beta) = (0, 0)$: $T_{2 \rightarrow 1} = -1.37 \times 10^{-5}$, $T_{1 \rightarrow 2} = 1.28 \times 10^{-5}$.

(The results may differ slightly with different series due to the random number generation.) For the case $\alpha = 0.5$, $|T_{2 \rightarrow 1}/T_{1 \rightarrow 2}| > 740$, one may then conclude that this is a one-way causality from X_2 to X_1 , as is indeed true. For the latter, however, one actually cannot say much from the numbers. Although they are small, they tell no more than that the information flows in both directions are of equal importance. From this particular example, one sees that, though theoretically the information flows in both directions should be precisely zero, in reality the rates evaluated from two time series, albeit very small, generally do not precisely vanish. One then cannot tell whether the causality indeed exists. We need to normalize the obtained information flow for one to see the relative magnitude.

Of course, one may perform statistical test for the results. In the first example, at a 90% significance level, $T_{2 \rightarrow 1} = 0.1481 \pm 0.0015$, $T_{1 \rightarrow 2} = -0.0002 \pm 0.0015$, so $T_{1 \rightarrow 2}$ is insignificant. For the second example, $T_{2 \rightarrow 1} = -1.37 \times 10^{-5} \pm 2.19 \times 10^{-5}$, $T_{1 \rightarrow 2} = 1.28 \times 10^{-5} \pm 2.19 \times 10^{-5}$. That is to say, at a 90% level, these flow rates are not significantly different from zero. In this sense, it seems that we could indeed infer rather accurately the true causality. However, a statistical test just tells how precise the estimate is; it does not tell how the information flow may weigh in the entropy balance of the series. Besides, it depends on the length of the series which is irrelevant to the parameter to be estimated. To see this more clearly, consider the following case:

$$\alpha = 0.01, \quad \beta = 0.01.$$

Obviously, the information flows, albeit existing, make only tiny contributions to their respective series, as the coupling coefficients are over an order smaller; in classical perturbation analysis, they can be dropped to the first order approximation. The computed information flows, at a 90% significance level, are $T_{2 \rightarrow 1} = (1.648 \pm 0.692) \times 10^{-4}$, $T_{1 \rightarrow 2} = (1.123 \pm 0.639) \times 10^{-4}$. These results indicate that they are significantly different from zero—This from one aspect testifies to the success of the formalism. However, the small numbers cannot tell how important they are, since, with a slowly varying series, even the dominant flow could be very small. On the other hand, if we cut the series by half and pick the first 40000 points for the analysis, then the result will be $T_{2 \rightarrow 1} = (0.653 \pm 0.751) \times 10^{-4}$, $T_{1 \rightarrow 2} = (1.240 \pm 0.690) \times 10^{-4}$. So one finds that $T_{2 \rightarrow 1}$ is insignificant while $T_{1 \rightarrow 2}$ is. (Again, these small numbers may show large fluctuation if series of different lengths are used, because in reality they are insignificant.) Can one thus conclude that there is a one-way causality, or can he/she thus asserts that this shortened series yields a more reliable estimation? Surely this is absurd. The problem here is that we do need a normalized flow to evaluate its importance relative to other factors.

The normalization is by no means as simple as it appears with correlation analysis, which is in an inner product form based on the Cauchy-Schwarz inequality. In the following we will see that we need to get down to the fundamentals of information flow before arriving at a logically and physically sound normalizer. As what we did before in [11], this normalizer is estimated with the method of maximal likelihood estimation, using the given time series (section III). The resulting formula is then validated (section IV) with the autoregression

example as shown above. To demonstrate its diverse applicability, presented subsequently are two real world examples, one from climate science (section V A), another from financial economics (section V B). This study is summarized in section VI.

II. INFORMATION FLOW NORMALIZATION

As mentioned in the introduction, the normalization is not as simple as it seems to be. A natural normalizer that comes to mind, at the hint of correlation coefficient, might be the information of a series transferred from itself. A snag is, however, that this quantity may turn out to be zero, just as that in the Hénon map, a benchmark problem we have examined before (see the references in [9]).

Another snag is, the two way causality actually cannot be normalized together, as that in correlation analysis based on the Cauchy-Schwarz inequality. That is to say, two information flows of equal size may have different relative importances in their respective series.

Since in our framework the information flow from X_2 to X_1 is the contribution of X_2 that makes to the time rate of change of the marginal entropy of X_1 , written H_1 , one may ask whether the rate of marginal entropy change dH_1/dt can be the normalizer. This might be appealing, but there is a third snag. As information flow can be positive or negative, dH_1/dt may turn out to be smaller than the flow in absolute value—The so-obtained relative flow would exceed 100%, a case which we do not want to see.

All the above tells that information flow normalization is by no means a trivial task. We need to get to the basics and analyze how an information flow within a system is derived. By Ref. [10], the time rate of change of the marginal entropy of X_1 is

$$\frac{dH_1}{dt} = -E \left(F_1 \frac{\partial \log \rho_1}{\partial x_1} \right) - \frac{1}{2} E \left(g_{11} \frac{\partial^2 \log \rho_1}{\partial x_1^2} \right). \quad (6)$$

It is actually a result of two mutually exclusive mechanisms: the first is the information flow $T_{2 \rightarrow 1}$ as shown in (2); the second is the complement, i.e., the rate of entropy increase without taking into account of the effect of X_2 . Denote this latter as $\frac{dH_{1\mathfrak{I}}}{dt}$, Liang (2008)[10] has proved that

$$\frac{dH_{1\mathfrak{I}}}{dt} = E \left(\frac{\partial F_1}{\partial x_1} \right) - \frac{1}{2} E \left(g_{11} \frac{\partial^2 \log \rho_1}{\partial x_1^2} \right) - \frac{1}{2} E \left(\frac{1}{\rho_1} \frac{\partial^2 g_{11} \rho_1}{\partial x_1^2} \right). \quad (7)$$

The right hand side has three terms. The first term is precisely the time rate of change of H_1 due to X_1 itself in the absence of stochasticity. This is the starting point which we have shown in 2005[8] in establishing the rigorous formalism and proved later on (cf. [9]). Hence through a careful analysis, the increase in the marginal entropy H_1 is decomposed into three parts:

$$\frac{dH_1^*}{dt} = E \left(\frac{\partial F_1}{\partial x_1} \right), \quad (8)$$

$$\frac{dH_{1\text{noise}}}{dt} = -\frac{1}{2} E \left(g_{11} \frac{\partial^2 \log \rho_1}{\partial x_1^2} \right) - \frac{1}{2} E \left(\frac{1}{\rho_1} \frac{\partial^2 g_{11} \rho_1}{\partial x_1^2} \right), \quad (9)$$

as well as $T_{2 \rightarrow 1}$ in (2), which correspond to, respectively, the contribution due to X_1 itself, the stochastic effect, and the information flowing from X_2 . Note this decomposition does

not appear explicitly in the marginal entropy evolution equation (6), as the two stochastic terms cancel out.

The normalization is now made easy. Let

$$Z \equiv |T_{2 \rightarrow 1}| + \left| \frac{dH_1^*}{dt} \right| + \left| \frac{dH_1^{noise}}{dt} \right|. \quad (10)$$

Obviously it is no less than $T_{2 \rightarrow 1}$ in magnitude, and cannot be zero unless X_1 does not change, a situation that is excluded in time series analysis. We may therefore pick Z as the normalizer, and define

$$\tau_{2 \rightarrow 1} = T_{2 \rightarrow 1}/Z. \quad (11)$$

This way if $\tau_{2 \rightarrow 1} = 1$, the variation of H_1 is 100% due to the information flow from X_2 ; if $\tau_{2 \rightarrow 1}$ is approximately zero, X_2 is not causal. Therefore, $\tau_{2 \rightarrow 1}$ assesses the importance of the influence of X_2 to X_1 relative to other processes.

It should be pointed out that, the above normalizer applies to $T_{2 \rightarrow 1}$ only. For $T_{1 \rightarrow 2}$, it is

$$|T_{1 \rightarrow 2}| + \left| \frac{dH_2^*}{dt} \right| + \left| \frac{dH_2^{noise}}{dt} \right|,$$

which may be quite different in value. This from another aspect reflects the asymmetry between $T_{2 \rightarrow 1}$ and $T_{1 \rightarrow 2}$.

III. ESTIMATION

As in Ref. [11], consider a linear version of the stochastic differential equation (SDE) (1)

$$d\mathbf{X} = \mathbf{f} + \mathbf{A}\mathbf{X}dt + \mathbf{B}d\mathbf{W}, \quad (12)$$

where \mathbf{f} is a constant vector, and $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ are constant matrices. Initially if \mathbf{X} obeys a Gaussian distribution, then it is a Gaussian for ever, i.e.,

$$\rho(\mathbf{x}) = \frac{1}{2\pi\sqrt{\det \Sigma}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \quad (13)$$

with the mean $\boldsymbol{\mu}$ and covariance matrix Σ governed by equations

$$\frac{d\boldsymbol{\mu}}{dt} = \mathbf{f} + \mathbf{A}\boldsymbol{\mu}, \quad (14)$$

$$\frac{d\Sigma}{dt} = \mathbf{A}\Sigma + \Sigma\mathbf{A}^T + \mathbf{B}\mathbf{B}^T. \quad (15)$$

So Eqs. (8) and (9) can be explicitly evaluated:

$$\frac{dH_1^*}{dt} = E(a_{11}) = a_{11}, \quad (16)$$

and

$$\frac{dH_1^{noise}}{dt} = -\frac{1}{2}E\left(g_{11}\frac{\partial^2 \log \rho_1}{\partial x_1^2}\right) - \frac{1}{2}E\left(\frac{1}{\rho_1}\frac{\partial^2 g_{11}\rho_1}{\partial x_1^2}\right)$$

$$\begin{aligned}
&= -\frac{1}{2}E \left[g_{11} \left(-\frac{1}{\sigma_1^2} \right) \right] - \frac{1}{2} \iint_{\mathbb{R}^2} \rho_{2|1} \frac{\partial^2 g_{11} \rho_1}{\partial x_1^2} dx_1 dx_2 \\
&= \frac{1}{2} \frac{g_{11}}{\sigma_1^2} - \frac{1}{2} \int_{\mathbb{R}} \frac{\partial^2 g_{11} \rho_1}{\partial x_1^2} \left(\int_{\mathbb{R}} \rho_{2|1}(x_2|x_1) dx_2 \right) dx_1,
\end{aligned}$$

since neither g_{11} nor ρ_1 depends on x_2 . But $\int_{\mathbb{R}} \rho_{2|1} = 1$, and ρ_1 is compactly supported, the whole second term on the right hand side then vanishes. Hence

$$\frac{dH_1^{noise}}{dt} = \frac{1}{2} \frac{g_{11}}{\sigma_{11}}, \quad (17)$$

where for notational symmetry σ_1^2 has been written as σ_{11} . These, together with the information flow from X_2 to X_1 as we have obtained before[8][11], $T_{2 \rightarrow 1} = \frac{\sigma_{12}}{\sigma_{11}} a_{12}$, form the three constituents that account for the evolution of the marginal entropy of X_1 .

An observation about $\frac{dH_1^{noise}}{dt} = g_{11}/(2\sigma_{11})$, where $g_{11} = b_{11}^2 + b_{12}^2$, is that it is always positive. That is to say, the noise always contributes to increase the marginal entropy of X_1 , conforming to our common sense. In financial economics, this reflects the volatility of, say, a stock. On the other hand, for a stationary series, i.e., when $d/dt = 0$, the balance on the right hand side of Eq. (15) requires that $2\sigma_{11} \sim g_{11}$. So this quantity is also related to the noise-to-signal ratio.

The above results need to be estimated if what we are given are just a pair of time series. That is to say, what we know is a single realization of some unknown system, which, if known, can produce infinitely many realizations. The problem now is turned into estimating (16) and (17) with the available statistics of the given time series.

We use maximum likelihood estimation (e.g., [12]) to achieve the goal. The procedure follows precisely that of [11], which for easy reference we briefly summarize here. As established before, a further assumption that $b_{12} = 0$, and hence $g_{11} = b_{11}^2$, will much simplify the result, while in practice this is quite reasonable.

Suppose that the series are equal-distanced with a time stepsize Δt , and let N be the sample size. Consider an interval $[n\Delta t, (n+1)\Delta t]$, and let the transition probability function (pdf) be $\rho(\mathbf{X}_{n+1}|\mathbf{X}_n; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ stands for the vector of parameters to be estimated. So the log likelihood is

$$\ell_N(\boldsymbol{\theta}) = \sum_{n=1}^N \log \rho(\mathbf{X}_{n+1}|\mathbf{X}_n; \boldsymbol{\theta}) + \log \rho(\mathbf{X}_1).$$

As N is usually large, the term $\rho(\mathbf{X}_1)$ can be dropped without causing much error. The transition pdf is, with the Euler-Bernstein approximation (see [11]),

$$\rho(\mathbf{X}_{n+1} = \mathbf{x}_{n+1} | \mathbf{X}_n = \mathbf{x}_n) = \frac{1}{[(2\pi)^2 \det(\mathbf{B}\mathbf{B}^T \Delta t)]^{1/2}} \times e^{-\frac{1}{2}(\mathbf{x}_{n+1} - \mathbf{x}_n - \mathbf{F}\Delta t)^T (\mathbf{B}\mathbf{B}^T \Delta t)^{-1} (\mathbf{x}_{n+1} - \mathbf{x}_n - \mathbf{F}\Delta t)},$$

where $\mathbf{F} = \mathbf{f} + \mathbf{A}\mathbf{X}$. This results in a log likelihood functional

$$\ell_N(\mathbf{f}, \mathbf{A}, \mathbf{B}) = \text{const} - \frac{N}{2} \log g_{11} g_{22} - \frac{\Delta t}{2} \left(\frac{1}{g_{11}} \sum_{n=1}^N R_{1,n}^2 + \frac{1}{g_{22}} \sum_{n=1}^N R_{2,n}^2 \right), \quad (18)$$

where

$$R_{i,n} = \dot{X}_{i,n} - (f_i + a_{i1}X_{1,n} + a_{i2}X_{2,n}), \quad i = 1, 2, \quad (19)$$

and $\dot{X}_{i,n}$ is the Euler forward differencing approximation of $\frac{dX_i}{dt}$:

$$\dot{X}_{i,n} = \frac{X_{i,n+k} - X_{i,n}}{k\Delta t}, \quad (20)$$

with $k \geq 1$. Usually $k = 1$ should be used to ensure accuracy, but in some cases of deterministic chaos and the sampling is at the highest resolution, one needs to choose $k = 2$. Maximizing ℓ_N , we find[11] that the maximizer $(\hat{f}_1, \hat{a}_{11}, \hat{a}_{12})$ satisfies the following algebraic equation:

$$\begin{bmatrix} \frac{1}{\bar{X}_1} & \frac{\bar{X}_1}{\bar{X}_1^2} & \frac{\bar{X}_2}{\bar{X}_1\bar{X}_2} \\ \frac{\bar{X}_1}{\bar{X}_2} & \frac{\bar{X}_1\bar{X}_2}{\bar{X}_2^2} & \frac{\bar{X}_2}{\bar{X}_2^2} \end{bmatrix} \begin{bmatrix} \hat{f}_1 \\ \hat{a}_{11} \\ \hat{a}_{12} \end{bmatrix} = \begin{bmatrix} \frac{\bar{\dot{X}}_1}{\bar{X}_1\bar{\dot{X}}_1} \\ \frac{\bar{\dot{X}}_1}{\bar{X}_2\bar{\dot{X}}_1} \end{bmatrix}, \quad (21)$$

where the overline signifies sample mean. After some manipulations (see [11]), this yields the MLE estimators:

$$\hat{a}_{11} = \frac{C_{22}C_{1,d1} - C_{12}C_{2,d1}}{\det \mathbf{C}}, \quad (22)$$

$$\hat{a}_{12} = \frac{-C_{12}C_{1,d1} + C_{11}C_{2,d1}}{\det \mathbf{C}}, \quad (23)$$

$$\hat{g}_{11} = \frac{Q_{N,1}\Delta t}{N}, \quad (24)$$

$$\hat{f}_1 = \bar{\dot{X}}_1 - \hat{a}_{11}\bar{X}_1 - \hat{a}_{12}\bar{X}_2, \quad (25)$$

where

$$C_{ij} = \overline{(X_i - \bar{X}_i)(X_j - \bar{X}_j)}, \quad (26)$$

$$C_{i,dj} = \overline{(X_i - \bar{X}_i)(\dot{X}_j - \bar{\dot{X}}_j)}, \quad (27)$$

are the sample covariances, and

$$\begin{aligned} Q_{N,1} &= \sum_{n=1}^N \left[\dot{X}_{1,n} - (\hat{f}_1 + \hat{a}_{11}X_{1,n} + \hat{a}_{12}X_{2,n}) \right]^2 \\ &= \sum_{n=1}^N \left[(\dot{X}_{1,n} - \bar{\dot{X}}_{1,n}) - \hat{a}_{11}(X_{1,n} - \bar{X}_1) - \hat{a}_{12}(X_{2,n} - \bar{X}_2) \right]^2 \\ &= N(C_{d1,d1} + \hat{a}_{11}^2 C_{11} + \hat{a}_{12}^2 C_{22} - 2\hat{a}_{11}C_{d1,1} - 2\hat{a}_{12}C_{d1,2} + 2\hat{a}_{11}\hat{a}_{12}C_{12}). \end{aligned} \quad (28)$$

On the other hand, the population covariance matrix Σ can be rather accurately estimated by the sample covariance matrix \mathbf{C} . So (16)-(17) become

$$\frac{dH_1^*}{dt} = \frac{C_{22}C_{1,d1} - C_{12}C_{2,d1}}{C_{11}C_{22} - C_{12}^2}, \quad (29)$$

$$\frac{dH_1^{noise}}{dt} = \frac{\Delta t}{2NC_{11}}Q_{N,1} \quad (30)$$

As that in Ref. [11] with $T_{2 \rightarrow 1}$, here $\frac{dH_1^*}{dt}$ and $\frac{dH_1^{noise}}{dt}$ should bear a hat, since they are the corresponding estimators. We abuse the notation a little bit to avoid notational complexity;

from now on they should be understood as their respective estimators. With these the normalizer is

$$Z = |T_{2 \rightarrow 1}| + \left| \frac{dH_1^*}{dt} \right| + \left| \frac{dH_1^{noise}}{dt} \right|, \quad (31)$$

and hence we have the relative information flow from X_2 to X_1 :

$$\tau_{2 \rightarrow 1} = \frac{T_{2 \rightarrow 1}}{Z}. \quad (32)$$

IV. THE AUTOREGRESSIVE EXAMPLE REVISITED

Back to the autoregressive process exemplified in the beginning. When $\alpha = 0$, $\beta = 0$, the computed relative information flow rates are:

$$\tau_{2 \rightarrow 1} = -0.0016\%, \quad \tau_{1 \rightarrow 2} = 0.0018\%.$$

Clearly both are negligible in comparison to the contributions from the other processes in their respective series. This is in agreement with what one would conclude based on the absolute information flow computation and statistical testing. For the case $\alpha = \beta = 0.01$, in which one may encounter difficulty due to the ambiguous small numbers, the computed relative information flow rates are:

$$\tau_{2 \rightarrow 1} = 0.018\%, \quad \tau_{1 \rightarrow 2} = 0.015\%.$$

Again they are essentially negligible, just as one would expect.

On the other hand, when $\alpha = 0.5$, $\beta = 0$,

$$\tau_{2 \rightarrow 1} = 17\%, \quad \tau_{1 \rightarrow 2} = -0.03\%.$$

To X_1 , the influence from X_2 is large, contributing to more than 1/6 of the total entropy change. In contrast, the influence from X_1 to X_2 is negligible.

It should be pointed out that the relative information flow, say, $\tau_{2 \rightarrow 1}$, makes sense only with respect to X_1 , since the comparison is within the series itself. Here comes the following situation: For a two-way causal system with absolute information flows $T_{2 \rightarrow 1}$ and $T_{1 \rightarrow 2}$ of equal importance, their relative importances within their respective series could be quite different. For example,

$$\begin{aligned} X_1(n+1) &= -0.5X_1(n) + 0.9X_2(n) + 2e_1(n), \\ X_2(n+1) &= -0.2X_1(n) + 0.5X_2(n) + e_2(n). \end{aligned}$$

where $e_1(n)$ and $e_2(n)$ are identical independent normals ($\sim N(0,1)$). Initialize them with random values between $[0,1]$ and generate 80000 data points on MATLAB. The computed information flow rates (in nats per iteration)

$$|T_{2 \rightarrow 1}| = 0.13, \quad |T_{1 \rightarrow 2}| = 0.12,$$

which are almost the same. The relative information flows, however, are quite different:

$$|\tau_{2 \rightarrow 1}| = 6.7\%, \quad |\tau_{1 \rightarrow 2}| = 13\%.$$

In terms of relative contribution in their respective series, the former is way more below the latter.

Generally speaking, the above imbalance is a rule, not an exception, reflecting the asymmetry of information flow. One may reasonably imagine that, in some extreme situation, a flow might be dominant while its counterpart is negligible within their respective series, although the two are of the same order in absolute value.

V. APPLICATIONS

A. Re-examination of the ENSO-IOD relationship

El Niño, also known as El Niño-Southern Oscillation, or ENSO for short, is a long known and extensively studied climate mode in the tropical Pacific Ocean due to its relation to the global disasters like the droughts in Southeast Asia, Southern Africa, and northern Australia, the floods in Ecuador, the increasing number of Typhoons, the death of birds and dolphins in Peru, and the famine and epidemic diseases in far-flung regions of the world[13]. A correct forecast of an El Niño (or its cold counterpart, La Niña) a few months earlier will not only help issue in-advance warnings of potential disastrous impacts, but also make the subsequent seasonal weather forecasting much easier. However, this aperiodic leading mode in the tropical Pacific seems to be extraordinarily difficult to predict. A good example is the latest “Super El Niño ” or “Monster El Niño ”, which has been predicted to arrive in 2014 in a lot of portentous forecasts, turns out to be a computer artifact.

For more reliable predictions, it is imperative to clarify the source of its uncertainty or unpredictability. In Ref. [11], we have presented an application of Eq. (3) to the relation study between El Niño and the Indian Ocean Dipole (IOD), another major climate mode in the Indian Ocean[14], and found that the Indian Ocean is a source of uncertainty that keeps ENSO from being accurately predicted. Since in that study there is no relative importance assessment, we do not know whether the information flows, albeit significant, do weigh much in the modal variabilities. We hence redo the computation using the relative information flow formula (32).

We use for this study the same data as that used in [11], which include the Niño4 index series and the sea surface temperature (SST) series downloaded from the NOAA ESRL Physical Sciences Division[15], and the IOD index namely DMI series from the JAMSTEC site[16].

Shown in Fig. 2 is the relative information flow rate from the Indian Ocean SST to El Niño, $\tau_{IO \rightarrow ENSO}$. From it one can see that the information flow accounts for more than 10% of the uncertainties of Niño4, the maximum reaching 27%. This number is very large. Besides, all the values are positive, indicating that the Indian Ocean SST functions to make El Niño more uncertain. No wonder recently researchers find that assimilation of the Indian Ocean data helps the prediction of El Niño (e.g., [17]), although traditionally the Indian Ocean is mostly ignored in El Niño modeling.

Besides the relative importance we have just obtained, Fig. 2 also reveals some difference in structure from its counterpart, i.e., the Fig. 5b of [11].[22] A conspicuous difference is that now there are clearly two centers, residing on either side of Indian. Note this structure is different from the traditional dipolar pattern as one would expect; here both centers are positive. This means that the Northern Indian Ocean SST anomalies, both the positive phase and negative phase, as an integral entity influence the El Niño variabilities, and,

in particular, make El Niño more unpredictable. The dipolar structure implies that most probably this entity is IOD, not others like IOBM (Indian Ocean Basin Mode; see [18]).

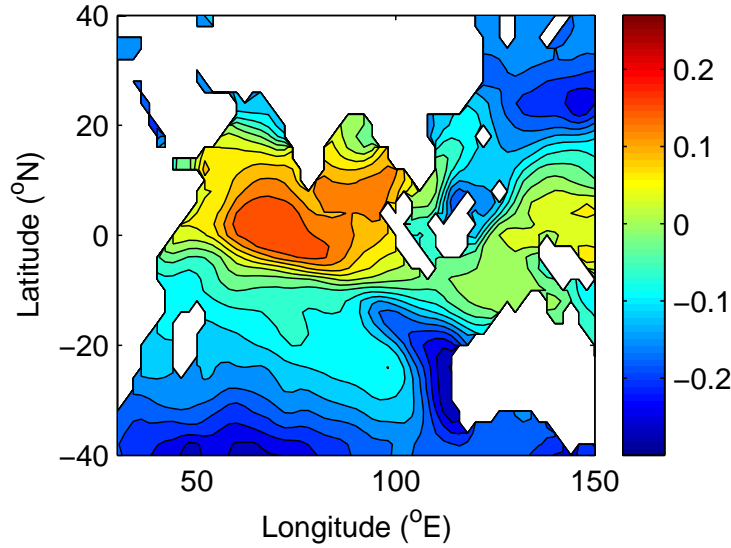


FIG. 2: (Color online) Relative information flow from the Indian Ocean SST to Niño4, $\tau_{IO \rightarrow ENSO}$.

To see more about this, we look at the information flow from the index DMI to the tropic Pacific SST. The absolute rates are referred to the Fig. 4a of [11]; shown in Fig. 3 are $\tau_{DMI \rightarrow Pacific}$. Indeed the computed flow rates are significant, and all are positive. The largest τ , which occupies a large swathe of the equatorial region between 175°W through 135°W, reaches 10%. Moreover, the structure reminds one of the El Niño pattern. It is generally the same as that in its counter part, i.e., Fig. 4a of [11], save for two changes: (1) the maximum center moves westward; (2) the small center of a secondary maximum near 125°E at the equator disappears. This clear El Niño-like structure attests to the above conjecture that IOD is indeed a major source of uncertainty for the El Niño forecast.

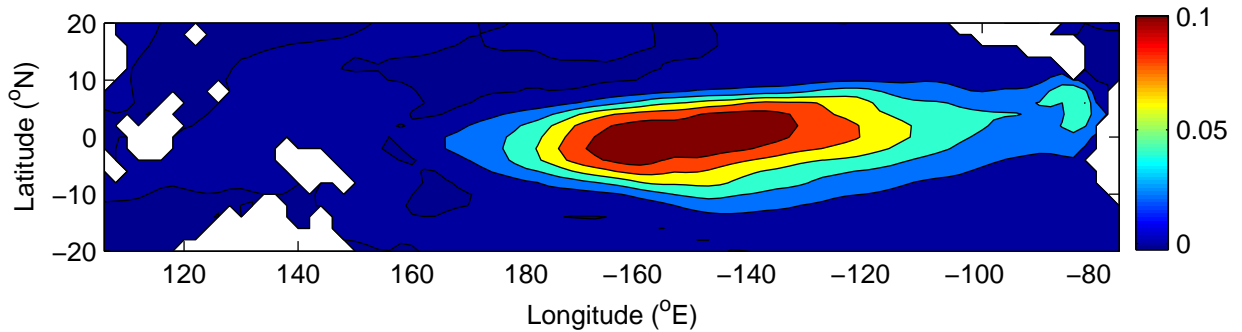


FIG. 3: (Color online) Relative information flow from the IOD index to the tropical Pacific SST, $\tau_{IOD \rightarrow Pacific}$.

We have also computed the relative information flows from El Niño to the Indian Ocean

SST, and that from the Pacific SST to the IOD, using the same datasets. The results are also significant, though only approximately half as shown above.

B. A financial example

We now look at the causal relations between several financial time series. Here it is not our intention to conduct an financial economics research or study the market dynamics from an econophysical point of view; our purpose is to demonstrate a brief application of the aforementioned formalism for time series analysis. Nonetheless, this topic is indeed of interest to both physicists and economists in the field of macroscopic econophysics; see, for example, [19].

We pick nine stocks in the United States and download their daily prices from ^{YAHOO!}_{finance}. These stocks are: MSFT (Microsoft Corporation), AAPL (Apple Inc.), IBM (International Business Machines Corporation), INTC (Intel Corporation), GE (General Electric Company), WMT (Wal-Mart Stores Inc.), XOM (Exxon Mobil Corporation), CVS (CVS Health Corporation), F (Ford Motor Corporation). Among these are high-tech companies (MSFT, AAPL, IBM, INTC), retail trade companies [e.g., the drugstore chains (CVS) and discount stores (WMT)], automotive industry (F), oil and gas industry (XOM), and the multinational conglomerate corporation GE which operates through the segments of energy, technology infrastructure, capital finance, etc. Here by “daily” we mean on a trading day basis, excluding, say, holidays and weekends. Since stock prices are generally nonstationary, we check the series of daily return, i.e.,

$$R(t) = [P(t + \Delta t) - P(t)]/P(t),$$

or log-return

$$r(t) = \ln P(t + \Delta t) - \ln P(t),$$

where $P(t)$ are the adjusted closing prices in the yahoo spreadsheet, and Δt is one trading day. Following most people we use the series of log-returns r for our purpose. In fact, return and log-return series are approximately equivalent, particularly in the high-frequency regime, as indicated by [20]. Since the most recent stock MSFT started on March 13, 1986, all the series are chosen from that date through December 26, 2014, when we started to examine these series. This amounts to 7260 data points, and hence 7259 points for the log-return series.

Using Eq. (3), we compute the information flows between the nine stocks and form a matrix of flow rates; see Table I. A flow direction is represented with the matrix indices; more specifically, it is from the row index to the column index. For example, listed at the location (2,4) is $T_{2 \rightarrow 4}$, i.e., $T_{AAPL \rightarrow INTC}$, the flow rate from Apple to Intel, while (4,2) stores the rate of the reverse flow, $T_{INTC \rightarrow AAPL}$. Also listed in the table are the respective confidence intervals at the 90% level.

From Table I, most of the information flow rates are significant at the 90% level, as highlighted. Their values vary from 4 to 22 (units: 10^{-3} nats/day; same below in this section). The maximum is $|T_{IBM \rightarrow XOM}| = 22$, and second to it are $|T_{WMT \rightarrow CVS}|$ and $|T_{CVS \rightarrow GE}|$, both being 21. The influence of IBM to Exxon is not a surprise, considering the dependence of the oil industry on high-tech equipments. The mutual causality between the retail stores WMT and CVS are also understandable. The information flow from CVS to GE could be through the sales of GE products; after all, GE makes household appliances. For the rest in the table, they can be summarized from the following two aspects.

(1) Companies as sources.

Look at the table row by row. Perhaps the most conspicuous feature is that the whole CVS row is significant. Next to it is XOM, with only three entries insignificant. That is to say, CVS has been found causal to all other stocks, though the causality magnitudes are yet to be assessed (see below). This does make sense. As a chain of convenience stores, CVS connects most of the general consumers and commodities and hence the corresponding industries. For XOM, it is also understandable that why it makes a source of causality. Oil or gas is for sure one of the most fundamental component in the American economy.

(2) Companies as recipients.

Examining column by column, the most outstanding stock is again CVS, with only one entry insignificant. That is to say, CVS is influenced by all other stocks except XOM. Following CVS is XOM, WMT, and INTC. The IBM and MSFT columns form the third tier.

A few words regarding the stock F. As a cause to other stocks (though causality maybe tiny), XOM has not been identified to be causal to F. In fact, F has not been found causal to XOM, either. This is a little surprising; the reason(s) can be found only after a careful analysis of Ford, which is beyond the scope here. (In fact, computation does reveal information flows between XOM and Toyota.) Interestingly, $|T_{F \rightarrow WMT}| > |T_{F \rightarrow CVS}|$. This is easy to understand, as we rely on our motor vehicles to shop at Wal-Mart, while CVS stores could be just somewhere in the neighborhood!

TABLE I: The rates of absolute information flow between the 9 chosen stocks (in 10^{-3} nats per trading day). At each entry the direction is from the row index to the column index of the matrix. Also listed are the standard errors at a 90% significance level, and highlighted are the significant flows.

	MSFT	AAPL	IBM	INTC	GE	WMT	XOM	CVS	F
MSFT	/	5±7	-3±8	-12±11	1±8	-1±6	-12±6	10±4	3±5
AAPL	-2±7	/	-11±7	-2±9	-7±6	-4±5	-11±4	4±3	-2±4
IBM	0±8	5±7	/	-9±9	-8±9	-11±6	-22±7	6±4	1±6
INTC	16±11	10±9	-7±9	/	0±8	-2±6	-12±5	11±4	3±6
GE	2±8	-3±6	-13±9	-16±8	/	-10±9	-6±9	14±6	6±9
WMT	10±6	7±5	4±6	-5±6	6±9	/	0±6	21±7	9±5
XOM	-10±6	-3±4	-14±7	-13±5	-15±9	-17±6	/	4±5	-1±6
CVS	-9±4	-5±3	-12±4	-11±4	-21±6	-17±7	-17±5	/	-7±4
F	0±5	0±4	0±6	-10±6	6±9	-13±5	1±6	6±4	/

The above significant absolute information flows, large or small, still need to be assessed regarding their respective relative importances before any conclusion of causality is reached. Using Eq. (32), we compute the relative information flow rates (in percentage) and tabulate them in Table II. The matrix entries are arranged in the same way as above. For clarity, those above or equal to 1% are highlighted. In contrast to Table I, we see only a few of the information flows account for more than 1% of their respective fluctuations. These include $|\tau_{IBM \rightarrow XOM}|$, $|\tau_{CVS \rightarrow GE}|$, $|\tau_{WMT \rightarrow CVS}|$, $|\tau_{CVS \rightarrow WMT}|$, $|\tau_{CVS \rightarrow XOM}|$, $|\tau_{XOM \rightarrow WMT}|$,

$|\tau_{INTC \rightarrow MSFT}|$, $|\tau_{GE \rightarrow INTC}|$, and $|\tau_{XOM \rightarrow GE}|$, the first three being the largest. This echoes what we have introduced in the beginning: though significant, some information flows may be negligible in their own marginal entropy balances.

TABLE II: As Table I, but for relative information flow (in percentage).

	MSFT	AAPL	IBM	INTC	GE	WMT	XOM	CVS	F
MSFT	/	0.3	-0.2	-0.8	0.0	0.0	-0.7	0.6	0.2
AAPL	-0.1	/	-0.7	-0.1	-0.5	-0.2	-0.7	0.2	-0.1
IBM	0.0	0.3	/	-0.6	-0.5	-0.7	-1.3	0.4	0.1
INTC	1.0	0.7	-0.4	/	0.0	-0.1	-0.7	0.7	0.2
GE	0.1	-0.2	-0.8	1.0	/	-0.6	-0.3	0.9	0.4
WMT	0.6	0.4	0.2	-0.3	0.4	/	0.0	1.3	0.6
XOM	-0.6	-0.2	-0.9	-0.9	-1.0	-1.1	/	0.3	-0.1
CVS	-0.6	-0.3	-0.8	-0.7	-1.3	-1.1	-1.1	/	-0.5
F	0.0	0.0	0.0	-0.6	0.4	-0.8	0.0	0.4	/

It should be noted that the causal relations generally change with time. If the series are long enough, we may look at how these information flows may vary from period to period. Pick the pair (IBM, GE) as an example. For the duration (March 1986 through present) considered above, $T_{GE \rightarrow IBM} = -13 \pm 9$, while $T_{IBM \rightarrow GE}$ is not significant. Neither $\tau_{GE \rightarrow IBM}$ nor $\tau_{IBM \rightarrow GE}$ reaches 1%. Since from the yahoo site both GE and IBM can be dated back to January 2, 1962, we can extend the time series a lot up to 13338 data points. Shown in Fig. 4a are the series of their historic prices, and in Fig. 4b and 4c are the corresponding log-returns.

Computation of information flows with the whole series (13338 points) results in $\tau_{IBM \rightarrow GE} = 1.6\%$ and $\tau_{GE \rightarrow IBM} = -0.5\%$, and $T_{IBM \rightarrow GE} = (27 \pm 6) \times 10^{-3}$, $T_{GE \rightarrow IBM} = (7 \pm 6) \times 10^{-3}$ nats/day, both being significant at the 90% level. This is very different from what are shown in Tables I and II, with the causal structure changed from a weak two-way causality to a stronger and more or less one-way causality. Since in the above only the data of the recent 30 years are used, we expect that in the early years this causal structure could be much enhanced. Choose the first 7000 points (from January 1962 through November 1989), the computed relative information flow rates are:

$$\begin{aligned}\tau_{IBM \rightarrow GE} &= 3.1\%, & \tau_{GE \rightarrow IBM} &= -0.2\%; \\ T_{IBM \rightarrow GE} &= 54 \pm 8, & T_{GE \rightarrow IBM} &= 3 \pm 8,\end{aligned}$$

where the units for the latter pair are in 10^{-3} nats/day, same below in this section. Further narrow down the period to 2250-3250 (corresponding to the period 1971-1975), then

$$\begin{aligned}\tau_{IBM \rightarrow GE} &= 5.7\%, & \tau_{GE \rightarrow IBM} &= -0.99\%; \\ T_{IBM \rightarrow GE} &= 101 \pm 21, & T_{GE \rightarrow IBM} &= 14 \pm 21,\end{aligned}$$

attaining the maximum of $\tau_{IBM \rightarrow GE}$, in contrast to the insignificant flow in Table I. Obviously, during this period, the causality can be viewed as one-way, i.e., from IBM to GE. And the relative flow makes more than 5%, much larger than those in Table II.

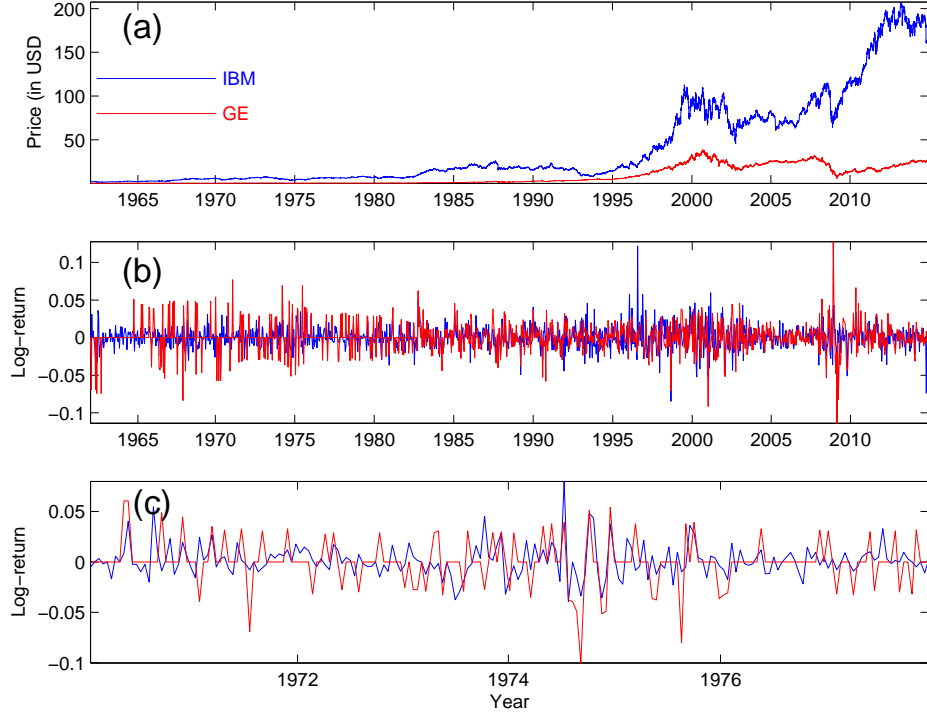


FIG. 4: (Color online) (a) Historic prices of IBM and GE (in US dollars). (b) Log-returns of the IBM and GE stocks. (c) A close-up of (b).

The above remarkable causal structure for that particular period actually can trace its reason back in the history of GE[21]. There is such a period in 1960's when "Seven Dwarfs" (Burroughs, Sperry Rand, Control Data, Honeywell, General Electric, RCA and NCR) competed with IBM the giant for computer business, and, particularly, to build mainframes. In 1965, GE had only a 3.7-percent market share of the industry, though it was then dubbed as the "King of the Dwarfs", while IBM had 65.3% share. Historically GE was once the largest computer user outside the US Federal Government; it got into computer manufacturing to avoid dependency on others. And, indeed, throughout the 60s, the causalities between GE and IBM are not significant. Then, why, as time entered 70s, was the information flow from IBM to GE suddenly increased to its highest level? It turned out that GE sold its computer division to Honeywell in 1970; in the following years (starting from 1971), it relied much on the IBM products. This GE computer era, which has almost gone to oblivion, does substantiate the existence of a causation between GE and IBM, and, to be more precise, an essentially one-way causation from IBM to GE. In this sense, our formalism is well validated.

VI. CONCLUSIONS

To assess the importance of a flow of information from a series, say, $\{X_2(n)\}$, to another, say, $\{X_1(n)\}$, it needs to be normalized. The normalization cannot follow a way as that in computing a correlation coefficient, since there is no such a theorem like the Cauchy-Schwarz inequality for it to base. Getting down to the fundamentals, we were able to distinguish three types of mechanisms that contribute to the evolution of the marginal entropy of X_1 ,

H_1 ,

- $\frac{dH_1^*}{dt}$: the contribution from X_1 its own;
- $T_{2 \rightarrow 1}$: the information flow from X_2 ;
- $\frac{dH_1^{noise}}{dt}$: the contribution from noise.

Similarly there are three such quantities for X_2 , as schematized in Fig. 5. We hence proposed that the normalization can be fulfilled as follows:

$$\tau_{2 \rightarrow 1} = \frac{T_{2 \rightarrow 1}}{\left| \frac{dH_1^*}{dt} \right| + \left| \frac{dH_1^{noise}}{dt} \right| + |T_{2 \rightarrow 1}|},$$

$$\tau_{1 \rightarrow 2} = \frac{T_{1 \rightarrow 2}}{\left| \frac{dH_2^*}{dt} \right| + \left| \frac{dH_2^{noise}}{dt} \right| + |T_{1 \rightarrow 2}|}.$$

Obviously, a normalized flow tells its importance relative to other mechanisms within its own series. In other words, the two flows are normalized differently, echoing the property of asymmetry which makes information flow analysis distinctly different from those such as correlation analysis or mutual information analysis.

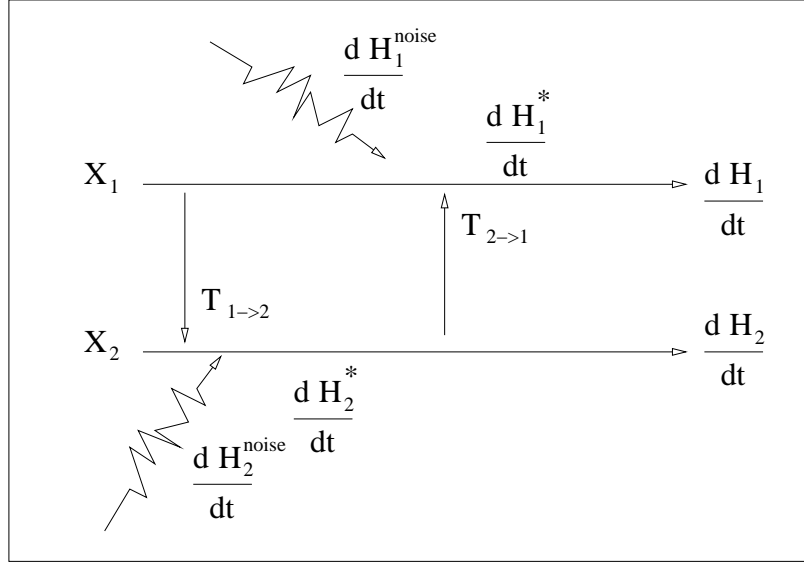


FIG. 5: A schematic of the marginal entropy evolutions and information flows in the system of (X_1, X_2) .

The above normalizer can be accurately obtained in the framework of a dynamical system. When only two equi-distanced series, say, X_1 and X_2 , are given, with a linear model its

constituents can be estimated as follows, in a nutshell,

$$\frac{dH_1^*}{dt} = p, \quad (33)$$

$$\frac{dH_1^{noise}}{dt} = \frac{\Delta t}{2C_{11}} [C_{d1,d1} + p^2 C_{11} + q^2 C_{22} - 2pC_{d1,1} - 2qC_{d1,2} + 2pqC_{12}], \quad (34)$$

$$T_{2 \rightarrow 1} = \frac{C_{11}C_{12}C_{2,d1} - C_{12}^2 C_{1,d1}}{C_{11}^2 C_{22} - C_{11}C_{12}^2}, \quad (35)$$

where Δt is the time stepsize,

$$p = \frac{C_{22}C_{1,d1} - C_{12}C_{2,d1}}{C_{11}C_{22} - C_{12}^2},$$

$$q = \frac{-C_{12}C_{1,d1} + C_{11}C_{2,d1}}{C_{11}C_{22} - C_{12}^2},$$

C_{ij} the sample covariance between X_i and X_j , $C_{i,dj}$ the sample covariance between X_i and \dot{X}_j , and

$$\dot{X}_{i,n} = \frac{X_{i,n+k} - X_{i,n}}{k\Delta t}$$

($k = 1$; but for chaotic series sampled at high resolution, $k = 2$ may be needed).

It should be noted that a relative information flow is for the comparison purpose within its own series. The two reverse flows between two series can only be compared in terms of absolute value, since they belong to different series. In this sense, absolute and relative information flows should be examined simultaneously. This is clarified in the schematic diagram in Fig. 5, and has been exemplified in the validations with two autoregressive processes. It is quite normal that two identical information flows may differ a lot in relative importance with respect to their own series, as testified in our realistic applications. In some extreme situation, a pair of equal flows may find one dominant but another negligible in their respective entropy balances.

Partly for demonstration and partly for verification, we have presented two applications. The first is a re-examination of the climate science problem previously studied in Ref. [11]. Considering the fadeout of the recent portentous predictions of a “super” or “monster” El Niño, we have particularly focused on the predictability of El Niño. Our result reconfirmed that the Indian Ocean SST is a source of uncertainty to the El Niño prediction. We further clarified that the information flow from the Indian Ocean is mainly through the Indian Ocean Dipole (IOD).

Another realistic problem we have examined regards the causation between a few randomly picked American stocks. It is shown that many flows (and hence causalities), though significant at a 90% level, their respective importances relative to other mechanisms are mostly negligible. The resulting matrices of absolute and relative information flows provide us a pattern of causality mostly understandable using our common sense. For example, Ford has a larger influence on Wal-Mart than on CVS because people rely on motor vehicles to shop at Wal-Mart, while CVS could be just somewhere within a walking distance. A particularly interesting case is that we have identified a strong one-way causality from IBM to GE during the early stage of these companies. This has revealed to us the story of “Seven Dwarfs” competing IBM the giant for computer market. In an era when this story

has almost gone to oblivion (one even cannot find it from GE’s website), and GE may have left us an impression that it never built any computers, let alone a series of mainframes, this finding is indeed remarkable.

Acknowledgments. This study was partially supported by Jiangsu Provincial Government through the “Specially-Appointed Professor Program” (Jiangsu Chair Professorship) to XSL, and by the National Science Foundation of China (NSFC) under Grant No. 41276032.

-
- [1] e.g., K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, J. Bhattacharya, *Physics Reports*, 441(1), 1-46 (2007); J. Pearl, *Causality: Models, Reasoning, and Inference*. MIT Press, Cambridge, MA, 2nd edition (2009); M. Lungarella, K. Ishiguro, Y. Kuniyoshi, N. Otsu, *International Journal of Bifurcation and Chaos*, 17(3), 903-921 (2007).
 - [2] C. Granger, *Econometrica*, 37, 424 (1969).
 - [3] J.A. Vastano and H.L. Swinney, *Phys. Rev. Lett.* **60**, 1773 (1988).
 - [4] T. Schreiber, *Phys. Rev. Lett.* **85**, 461 (2000).
 - [5] L. Barnett, A.B. Barrett, and A.K. Seth, *Phys. Rev. Lett.* **103**, 238701 (2009).
 - [6] J. Runge, J. Heitzig, N. Marwan, and J. Kurths, *Phys. Rev. E* 86, 061121 (2012).
 - [7] J. Sun and E. Bolt, *Physica D*, 267, 49-57 (2014).
 - [8] X.S. Liang, R. Kleeman, *Phys. Rev. Lett.* **95**, 244101 (2005);
 - [9] X.S. Liang, *Entropy* **15**, 327 (2013).
 - [10] X.S. Liang, *Phys. Rev. E* **78**, 031113 (2008).
 - [11] X.S. Liang, Unraveling the cause-effect relation between time series. *Phys. Rev. E* **90**, 052150 (2014).
 - [12] P.H. Garthwaite, I.T. Jolliffe, and B. Jones, *Statistical Inference*. (Prentice Hall, Hertfordshire, UK 1995).
 - [13] e.g., M.A. Cane, *Science* **222**, 1189 (1983); J.M. Wallace, E.M. Rasmusson, P.P. Mitchell, V.E. Kousky, E.S. Sarachik, H. von Storch, *J. Geophys. Res.* **103**, 14241 (1998).
 - [14] H.H. Saji, B.N. Goswami, P.N. Vinayachandran, and T. Yamagata, *Nature* **401**, 360 (1999).
 - [15] <http://www.esrl.noaa.gov/psd/>.
 - [16] <http://www.jamstec.go.jp/>.
 - [17] D. Chen, M.A. Cane, *J. Comput. Phys.* **227**, 3625 (2008).
 - [18] J.L. Yang, Q.Y. Liu, and Z. Liu: Linking observations of the Asian Monsoon to the Indian Ocean SST: Possible roles of Indian Ocean Basin Mode and Dipole Mode. *J. Clim.* 2010, 23(21), 5889-5902.
 - [19] R. Marshinski and H. Kantz, Analysing the information flow between financial time series: An improved estimator for transfer entropy. *Eur. Phys. J. B* **30**, 275-281 (2002).
 - [20] R. N. Mantegna, H.E. Stanley, *An Introduction to Econophysics. Correlations and complexity in finance* (Cambridge University Press, Cambridge, 2000)
 - [21] http://en.wikipedia.org/wiki/General_Electric/
 - [22] (Note: There is a typo in the caption of the Fig. 5 of [11]. The units should be nats/year, not nats/month.)